What is claimed is:

1. A method for identifying a biological sample, comprising:

generating a data set indicative of the composition of the biological sample;

denoising the data set to generate denoised data;

deleting the baseline from the denoised data to generate an intermediate data set;

defining putative peaks for the biological sample;

using the putative peaks to generate a residual baseline;

removing the residual baseline from the intermediate data set to generate a corrected data set;

locating, responsive to removing the residual baseline, a probable peak in the corrected data set; and

identifying, using the located probable peak, the biological sample.

2. The method according to claim 1 wherein the data set is a spectrometry data set.

3. The method according to claim 1 wherein the data set is generated by a mass spectrometer.

4. The method according to claim 1 wherein denoising the data set includes generating a noise profile for the data set.

5. The method according to claim 1 wherein denoising the data set includes

6. The method according to claim 5 further including generating a noise profile for stage 0.

7. The method according to claim 6 further including generating a noise profile for other stages.

8. The method according to claim 7 wherein the noise profile for each of the other stages is the noise profile for stage 0 scaled by a scaling factor.

9. The method according to clean 8 wherein the scaling factor is derived from the end portion of each of the other stages, respectively.

10. The method according to claim 5 further including applying a threshold to selected stages, the threshold being derived from the noise profile.

11. The method according to claim of 10 wherein the threshold is scaled by a threshold factor before being applied to the selected stages.

12. The method according to claim 7 wherein the threshold factor is selected so that higher stages of data are filtered less than lower stages.

13. The method according to claim 5 further including generating a sparse data set indicative of the denoised data.

14. The method according to claim 5 further including shifting the denoised data to account for variations due to a starting value for the wavelet transformation.

15. The method according to claim 1 wherein correcting the baseline further includes generating a moving average of the denoised data set.

16. The method according to claim 15 wherein the moving average is used to find peak sections in the denoised data set.

17. The method according to claim 16 wherein the peak sections are removed from the denoised data set.

18. The method according to claim 17 further including generating a baseline correction.

19. The method according to claim 1 further including compressing the intermediate data set, the intermediate data set having a plurality of data values associated with respective addresses.

20. The method according to claim 19 wherein a compressed data value is a real number that includes a whole portion representing the difference between two addresses.

21. The method according to claim 19 wherein a compressed data value is a real number that includes a decimal portion representing the difference between a maximum value of all the data values and a value at a particular address.

22. The method according to claim 1 further including performing a mass shift based on the position of the putative peaks.

23. The method according to claim 1 wherein generating the residual baseline includes deleting an area around each peak in the intermediate data.

24. The method according to claim 23 wherein the area deleted is derived from a determined width of a peak.

25. The method according to claim 23 wherein the residual baseline is derived from data remaining in the intermediate data after the peaks have been removed.

26. The method according to claim 23 wherein generating the residual. baseline includes fitting a quartic polynomial to the data remaining in the intermediate data after the peaks have been removed.

27. The method according to claim 1 wherein the probable peak is located by fitting a Gaussian curve to a peak area in the corrected data set.

28. The method according to claim 1 wherein the identifying step includes using a generated noise profile to calculate the signal-to-noise ratio for the probable peak.

29. The method according to claim 28 wherein a residual peak error is calculated by comparing the probable peak to a Gaussian curve.

30. The method according to claim 29 wherein the residual peak error is used to adjust the signal-to-noise ratio to generate an adjusted signal-to-noise ratio.

31. The method according to claim 1 wherein the identifying step includes deriving a peak probability for the probable peak.

32. The method according to claim 31 wherein the peak probability is derived using the signal-to-noise ratio.

33. The method according to claim 31 wherein the peak probability is derived by using an allelic ratio, the allelic ration being a comparison of two peak heights indicated in the corrected data.

34. The method according to claim 1 wherein the identifying step includes calculating a peak probability that a probable peak in the corrected data is a peak indicating composition of the biological sample.

35. The method according to claim 34 wherein peak probability is calculated for each of a plurality of probable peaks in the corrected data.

36. The method according to claim 35 wherein a highest probability is compared to a second-highest probability to generate a calling ratio.

37. The method according to claim 36 wherein the calling ratio is used to determine if the composition of the biological sample will be called.

38. A system for identifying a biological sample, the system comprising:

an instrument receiving the biological sample and generating a data set indicative of the composition of the biological sample;

a computer communicating to the instrument and configured to receive the generated data set, the computer performing the method of:

denoising the data set to generate denoised data;

deleting the baseline from the denoised data to generate an intermediate data set;

defining putative peaks for the biological sample;

removing the residual baseline from the intermediate data set to generate a corrected data set;

locating, responsive to removing the residual baseline, a probable peak in the corrected data set; and

identifying, using the located probable peak, the biological sample.

39. The system according to claim 38 wherein the computer is integral to the instrument.

40. A machine readable program operating on a computing device, the computing device being configured to receive a data set indicating composition of a biological sample, the program implement the steps of:

denoising the data set to generate denoised data;

deleting the baseline from the denoised data to generate an intermediate data set;

defining putative peaks for the biological sample;

using the putative peaks to generate a residual baseline;

removing the residual baseline from the intermediate data set to generate a corrected data set;

locating, responsive to removing the residual baseline, a probable peak in the corrected data set; and

identifying, using the located probable peak, the biological sample.

a mass spectrometer receiving the DNA sample and generating a data set indicative of the composition of the DNA sample;

a computing device configured to receive the data set, the computing device implementing the method comprising:

denoising the data set to generate denoised data;

removing sufficiently the baseline from the denoised data to generate a corrected data set;

locating a probable peak in the corrected data set; and

identifying, using the located probable peak, a component in the composition of the DNA sample.

42. The system according to claim 41, where the method further includes using a statistical methodology to determine if the located probable peak is an actual peak.

43. The system according to claim 41, where the method further includes determining whether the probability of the actual peak existing is sufficiently high to call the component of the DNA sample, and if the probability is not sufficiently high, then the method does not call the component.

44. The system according to claim 43, where the percentage of correctly called components is about 100 percent.

45. A system for identifying a component in a biological sample, comprising:

an instrument receiving the biological sample and generating a data set indicative of the component in the biological sample;

a computing device receiving the data set and performing the steps of:

generating corrected data by processing the data set to remove noise due to system and chemical reaction characteristics, the corrected data set having putative peak areas;

defining the position of expected peaks using known possible peak areas from the biological sample;

shifting the corrected data set to more closely align the putative peaks to the expected peaks;

calculating the probability that the putative peaks in the shifted data set are actual peaks;

calling the composition of the biological sample responsive to the calculated probability.